CLAIMS

What is claimed is:

1. A computer implemented method of crawling hyperlinked documents, comprising:

receiving a plurality of links to hyperlinked documents to be crawled;

grouping the plurality of links to hyperlinked documents by host;

selecting a host to crawl next according to a stall time of the host; and

crawling a hyperlinked document from the selected host.

2. The method of claim 1, wherein the stall time of the host is the earliest time in which a hyperlinked document from the host should be crawled.

3. The method of claim 1, wherein selecting a host to crawl next includes selecting a host with a stall time that is earlier than the current time.

4. The method of claim 1, further comprising grouping the hosts according to the number of hyperlinked documents to be crawled at each host.

5. The method of claim 4, further comprising examining the groups in descending order of the number of hyperlinked documents to be crawled at each host until a host is found with a stall time that is earlier than the current time.

6. The method of claim 4, wherein the hosts within each group are sorted by stall time.

7. The method of claim 4, further comprising moving the selected host to a group with one less hyperlinked documents to be crawled.

8. The method of claim 1, further comprising determining a retrieval time for retrieving the hyperlinked document from the selected host.

9. The method of claim 8, further comprising adjusting subsequent stall times for the selected host according to the retrieval times.

10. A computer program product for crawling hyperlinked documents, comprising:

computer code that receives a plurality of links to hyperlinked documents to be crawled;

computer code that groups the plurality of links to hyperlinked documents by host;

computer code that selects a host to crawl next according to a stall time of the host;

computer code that crawls a hyperlinked document from the selected host; and

a computer readable medium that stores the computer codes.

11. The computer program product of claim 10, wherein the computer readable medium is a CD-ROM, floppy disk, tape, flash memory, system memory, hard drive, or data signal embodied in a carrier wave.

12. A computer implemented method of crawling hyperlinked documents, comprising:

receiving a plurality of links to hyperlinked documents to be crawled;

grouping the plurality of links to hyperlinked documents by host;

selecting a host to crawl next according to a stall time of the host;

crawling a hyperlinked document from the selected host;

determining a retrieval time for retrieving the hyperlinked document from the selected host; and

adjusting subsequent stall times for the selected host according to the retrieval time.

13. The method of claim 12, wherein the stall time of the host is the earliest time in which a hyperlinked document from the host should be crawled.

14. The method of claim 12, wherein selecting a host to crawl next includes selecting a host with a stall time that is earlier than the current time.

15. The method of claim 12, further comprising grouping the hosts according to the number of hyperlinked documents to be crawled at each host.

16. The method of claim 15, further comprising examining the groups in descending order of the number of hyperlinked documents to be crawled at each host until a host is found with a stall time that is earlier than the current time.

17. The method of claim 15, wherein the hosts within each group are sorted by stall time.

18. The method of claim 15, further comprising moving the selected host to a group with one less hyperlinked documents to be crawled.

19. The method of claim 18, further comprising displaying the at least one category that was selected with the search results from the query.

20. A computer program product for crawling hyperlinked documents, comprising:

computer code that receives a plurality of links to hyperlinked documents to be crawled;

computer code that groups the plurality of links to hyperlinked documents by host;

computer code that selects a host to crawl next according to a stall time of the host;

computer code that crawls a hyperlinked document from the selected host including determining a retrieval time for retrieving the hyperlinked document from the selected host;

computer code that adjusts subsequent stall times for the selected host according to the retrieval time; and

a computer readable medium that stores the computer codes.

21. The computer program product of claim 20, wherein the computer readable medium is a CD-ROM, floppy disk, tape, flash memory, system memory, hard drive, or data signal embodied in a carrier wave.

22. A computer implemented method of crawling hyperlinked documents, comprising:

storing a plurality of links to hyperlinked documents to be crawled;

determining that more links to hyperlinked documents are desired;

sending requests to multiple link managers for more links to hyperlinked documents;

receiving additional links to hyperlinked documents from the link managers;

selecting a host to crawl next according to a stall time of the host; and

crawling a hyperlinked document from the selected host.

23. A computer program product for crawling hyperlinked documents, comprising:

computer code that stores a plurality of links to hyperlinked documents to be crawled;

computer code that determines that more links to hyperlinked documents are desired;

computer code that sends requests to multiple link managers for more links to hyperlinked documents;

computer code that receives additional links to hyperlinked documents from the link managers;

computer code that selects a host to crawl next according to a stall time of the host;

computer code that crawls a hyperlinked document from the selected host; and

a computer readable medium that stores the computer codes.

24. The computer program product of claim 23, wherein the computer readable medium is a CD-ROM, floppy disk, tape, flash memory, system memory, hard drive, or data signal embodied in a carrier wave.